



TITLE:

レゾルベントの多項式によるフィルタの伝達特性の調整 (数式処理とその周辺分野の研究)

AUTHOR(S):

村上, 弘

CITATION:

村上, 弘. レゾルベントの多項式によるフィルタの伝達特性の調整 (数式処理とその周辺分野の研究). 数理解析研究所講究録 2017, 2054: 168-181

ISSUE DATE:

2017-10

URL:

<http://hdl.handle.net/2433/237159>

RIGHT:

レゾルベントの多項式によるフィルタの伝達特性の調整

村上 弘

HIROSHI MURAKAMI

首都大学東京 数理情報科学専攻

DEPARTMENT OF MATHEMATICS AND INFORMATION SCIENCES,

Tokyo Metropolitan University *

要 約

与えられた実対称定値一般固有値問題の中間固有対で固有値が指定された区間にあるものをフィルタ対角化法を用いて求める。今回用いるフィルタは、シフトが虚数のレゾルベントの多項式の実部である。シフトと多項式の係数の値は、フィルタの伝達関数の特性がなるべく良いものとなるように調整する。

複数のレゾルベントの線形結合であるフィルタに比べると、レゾルベントの多項式の実部であるフィルタの伝達関数は達成可能な形状特性が劣る。しかし用いるレゾルベントは1つでよいので必要な記憶量を減らせる。また複数のレゾルベントを用いる場合と比べて行列分解に費やす演算量も減らせる。なぜならば、レゾルベントに対応する連立1次方程式を解くのに行列分解を用いる場合は、分解は最初に1度だけ行ない、同じ行列を再び分解する必要はないからである。

帯行列の固有値問題に対して数値実験を行ない、レゾルベントの多項式の実部をフィルタとして用いる方法は実際にうまく働くことを確認した。

ABSTRACT

For a given real symmetric-definite generalized eigenproblem, by the use of the filter diagonalization method we solve those internal eigenpairs whose eigenvalues are in a specified interval. In this paper, we use the filter operator which is the real-part of a polynomial of a resolvent with an imaginary shift. The shift of the resolvent and coefficients of the polynomial are tuned to make properties of the filter's transfer function well,

Compared from a filter which is a linear combination of many resolvents, a filter which is the real-part of a polynomial of a resolvent has poorer attainable properties of the transfer function. However, the amount of storage requirement is reduced since a single resolvent is used rather than many ones. The amount of computation required is also reduced. Because, when the simultaneous linear equations which corresponds to the resolvent is solved by some matrix factorization method, the matrix is factored only once, and there is no need to factor the same matrix again.

We made some experiments to solve banded eigenproblems, and found that the present method, which uses the real-part of a polynomial of a resolvent with an imaginary shift, worked well in practice.

*mrkmhrsh@tmu.ac.jp

1 はじめに

いま実対称定値一般固有値問題を $A\mathbf{v} = \lambda B\mathbf{v}$ とする。行列 A, B は実対称で、 B は正定値である。この問題の固有対で固有値が区間 $[a, b]$ の範囲にあるものをフィルタ対角化法を用いて解くことにする。まずこの固有値問題に対応するレゾルベントを $\mathcal{R}(\rho) = (A - \rho B)^{-1}B$ とする。求めたい固有対の固有値が固有値分布の中央にある（中間固有値）の場合には、レゾルベントが有界作用素になることを保証するために、シフト ρ を虚数にとる。そうして、適切に決めた虚数をシフトに持つレゾルベント $\mathcal{R}(\rho)$ の多項式の実部をフィルタ対角化法に用いるフィルタとして用いることにして、そのことについて考察を行なう。実部をとることで、フィルタは実作用素（実ベクトルから実ベクトルへの作用素）になる。

フィルタにレゾルベントの多項式の実部を用いる場合には、その伝達特性をなるべく望ましいものとするようにシフトと多項式の係数を調整して決定することが設計法になる。

しかし今回はフィルタおよびその伝達関数をチェビシェフ多項式を用いて表される形に制限することで、調整のための数値最適化の処理を省いて、設計を簡易に行なえるようにすることを試みた。

2 フィルタ作用素と伝達関数

実対称定値一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ の固有対であって固有値 λ が固有値分布の中間の指定された区間 $[a, b]$ にあるものを解くものとする。この固有値問題の固有値はすべて実数なので、固有値問題に対応するレゾルベントを $\mathcal{R}(\rho) = (A - \rho B)^{-1}B$ とするとき、シフト ρ を虚数に選ぶとレゾルベントは有界作用素になる。そうしてフィルタ \mathcal{F} として、レゾルベントの複素係数多項式の実部：

$$\mathcal{F} = c_{\infty}I + \operatorname{Re} \sum_{k=1}^n \gamma_k \{\mathcal{R}(\rho)\}^k$$

を採用する（ c_{∞} は実数、 γ_k , $k=1, 2, \dots, n$ は複素数である）。これは（実ベクトルから実ベクトルへの）実作用素である。そのとき、固有値問題の任意の固有対 (λ, \mathbf{v}) に対して $\mathcal{F}\mathbf{v} = f(\lambda)\mathbf{v}$ が成立する。ここで $f(\lambda)$ は、フィルタ \mathcal{F} による固有ベクトルの固有値だけで決まる伝達率を与える伝達関数であり、 $2n$ 次の実有理関数：

$$f(\lambda) = c_{\infty} + \operatorname{Re} \sum_{k=1}^n \frac{\gamma_k}{(\lambda - \rho)^k}$$

である。その極は虚数 ρ とその複素共役 $\bar{\rho}$ だけであり、どちらも n 位の極である。

行列 A, B が実対称でシフト ρ は虚数なので、レゾルベントの作用を与える連立 1 次方程式の係数 $C = A - \rho B$ は複素対称（ $C^T = C$ ）な正則行列になる。さらに、たとえば A, B が帯行列の場合には C も帯行列となるので、複素対称な帯行列 C を係数とする連立 1 次方程式は複素版の修正コレスキ法を用いることで能率よく解ける。（ピボット選択をしない修正コレスキ法を用いても結果の精度に支障がないかどうかについては、今後の検討を要する。もしも支障があるのならば対称性を利用せずに複素帯行列の連立 1 次方程式として片側ピボット選択を用いる帯 LU 分解を用いて解くことができるが、ピボット交換をしない修正コレスキ法と比べて必要な演算量と記憶量が増える。）注：修正コレスキ法では対称行列を LDL^T の形に分解するのに代数的な等式関係を四則演算だけを用いて解いているので、実対称行列に対する算法の実数の四則演算をそのまま複素数の四則演算に置き換えれば、複素対称行列に対する算法が得られる。

ベクトルにレゾルベントの n 次多項式を作用させる作業は同一のレゾルベントを適用する処理を n 回含むが、レゾルベントの作用を連立 1 次方程式を解いて与える際に行列の分解あるいは不完全分解を用いるときには、分解の結果が保持できるのならば分解は 1 度だけ行なえばよく、 n 回繰り返す必要はない。

指定された区間 $[a, b]$ に固有値 λ をもつ固有対を求める場合に、固有値の座標 λ に対してその正規化座標 t を $\lambda \in [a, b]$ から $t \in [-1, 1]$ への 1 次変換 $\lambda = \frac{a+b}{2} + \left(\frac{b-a}{2}\right)t$ により定義する。そうして正規化座標 t を引数とする伝達関数を $g(t) \equiv f(\lambda)$ で定義して、ある実数 $\mu > 1$ により、 $t \in [-1, 1]$ は通過域、 $\mu \leq |t|$ は阻止域、その中間の $1 < |t| < \mu$ は遷移域とする。伝達関数 $g(t)$ に対して課す制約条件は、1) 最大値が 1 であること、2) 通過域に於ける最小値は g_p で、逆に $g(t) \geq g_p$ となるのは通過域に限られること、3) さらに阻止域に於ける $|g(t)|$ の上限値が g_s であること、とする。レゾルベントのシフト ρ に虚数を用いる場合は、特に理由がなければ $g(t)$ を偶関数にするのが便利なので、以下では偶関数とする。そのとき互いに複素共役である $g(t)$ の極は純虚数になる。

- 遷移域と通過域の幅の形状比である μ の値は 1 より大きい方が望ましい。 μ が大きければ遷移域が広くなり、遷移域に多くの固有値が入る可能性が増す。固有値が遷移域にある固有対の数が増すと、フィルタで濾過するベクトルの数をそれだけ多くする必要が生じる。
- 阻止域に於ける伝達率の大きさの上限 g_s は極めて小さい正数であることが望ましい。 g_s が微小でなければ、固有値が阻止域にある不要な固有ベクトルが十分に減衰を受けずにフィルタで濾過された結果に混入して、それから張られる不変部分空間の近似を悪くする。
- 通過域での伝達率の最大最小比の逆数 g_p の値は 1 以下の正数であるが、1 に近いことが望ましい。 g_p の値が小さいと計算で求めた（固有値が通過域にある）近似対の精度はそれだけ不均一になる可能性がある。

ただし、これら $g(t)$ の形状に関係する μ , g_s , g_p の値の改良は互いに相反するので、適切なバランスを考えて最適に調整することになる。それはたとえば n を固定した場合に ρ と c_∞ と γ_k , $k = 1, 2, \dots, n$ の $n+2$ 個の変数に関する最適化計算となり、結果は数値としてだけ得られる。そこで以下では、チェビシェフ多項式の性質をうまく利用した設計法を示す。これは $g(t)$ の関数形を制限し、関数の選択の範囲を狭めていることになるので、 $n+2$ 個の変数の最適化の場合に比べると得られる伝達関数の形状の値は劣るはずであるが、制約条件を指定した場合の伝達関数 $g(t)$ を簡単な式で計算して決定できる利点がある。

3 伝達関数の設計

いま通過域は $t \in [-1, 1]$ 、阻止域は $t \in [\mu, \infty)$ とし、 σ を正の実数パラメタとして、チェビシェフ多項式を用いて伝達関数の関数形を以下の形に制限する。

$$g(t) = g_s T_n(y), \quad y = 2x - 1, \quad x = \frac{\sigma^2 + \mu^2}{\sigma^2 + t^2}.$$

このように表わされた正規化座標 t の伝達関数を決定するためのパラメタは n , μ , σ の 3 つである。阻止域の全体に対応してチェビシェフ多項式の引数 y は $(-1, 1]$ の全域を動く。この関数は t の $2n$ 次の実有理関数で、偶関数である。その極は複素共役な純虚数 $t = \pm \sigma\sqrt{-1}$ で、それぞれ n 位の極である。チェビシェフ多項式の性質により、阻止域 $|t| \geq \mu$ における制約条件 $|g(t)| \leq g_s$ は最初から満たされている。阻止域の外側 $t \in [0, \mu)$ では単調減少になる。残りの 2 つの制約条件 $g(0) = 1$ と $g(1) = g_p$ はそれぞれ、

$$\frac{1}{g_s} = T_n\left(1 + 2\frac{\mu^2}{\sigma^2}\right), \quad \frac{g_p}{g_s} = T_n\left(1 + 2\frac{\mu^2 - 1}{\sigma^2 + 1}\right)$$

となる。また逆双曲線関数を用いてこれらの制約条件を表わすと以下ようになる（公式 $\operatorname{arcosh}(1 + 2x^2) = 2 \operatorname{arsinh} |x|$ も用いている）。

$$\operatorname{arcosh} \frac{1}{g_s} = 2n \operatorname{arsinh} \frac{\mu}{\sigma}, \quad \operatorname{arcosh} \frac{g_p}{g_s} = 2n \operatorname{arsinh} \sqrt{\frac{\mu^2 - 1}{\sigma^2 + 1}}.$$

3.1 3つ組として n, μ, σ の値を指定する場合

いま3つ組 n, μ, σ の値を指定する場合は、チェビシェフ多項式あるいは(逆)双曲線関数を用いて表わされた以下の2つの関係式の右辺の式の値をそれぞれ計算すれば、ただちに g_s と g_p の値が求まる。

$$\frac{1}{g_s} = T_n \left(1 + 2 \frac{\mu^2}{\sigma^2} \right) = \cosh \left(2n \operatorname{arsinh} \frac{\mu}{\sigma} \right), \quad \frac{g_p}{g_s} = T_n \left(1 + 2 \frac{\mu^2 - 1}{\sigma^2 + 1} \right) = \cosh \left(2n \operatorname{arsinh} \sqrt{\frac{\mu^2 - 1}{\sigma^2 + 1}} \right)$$

もしも、得られた特性値 g_s と g_p の値が望ましくなければ、3つ組 n, μ, σ の値を変えて再試行する。

3.2 3つ組として g_p, g_s, μ の値を指定する場合

いま3つ組として伝達特性の形状の特徴的な3つの値である g_p, g_s, μ (ただし $1 > g_p \gg g_s > 0, \mu > 1$) を指定して、それらを満たすように伝達関数を表わすためのパラメタ σ と n の値を決定することを考える。上述した2つの制約条件の式から、 n の値を表わす式が2通り得られる。

$$n = \frac{\operatorname{arcosh} \frac{1}{g_s}}{2 \operatorname{arsinh} \frac{\mu}{\sigma}}, \quad n = \frac{\operatorname{arcosh} \frac{g_p}{g_s}}{2 \operatorname{arsinh} \sqrt{\frac{\mu^2 - 1}{\sigma^2 + 1}}}.$$

これらの2通りに表わされた n の式を等しいとおけば、 σ についての非線形方程式 $G(\sigma) = r$ が得られる。この方程式の左辺と右辺は具体的には以下のものである。

$$G(\sigma) = \frac{\operatorname{arsinh} \sqrt{\frac{\mu^2 - 1}{\sigma^2 + 1}}}{\operatorname{arsinh} \frac{\mu}{\sigma}}, \quad r = \frac{\operatorname{arcosh} \frac{g_p}{g_s}}{\operatorname{arcosh} \frac{1}{g_s}}.$$

これを解いて σ の値が得られたならば、上記の n の値を表わす2通りの式の値は一致していて、しかも n の値を表わすはずである。しかし n を表わす式を計算すると一般には実数値が得られて、 n が整数値であることに矛盾する。そこで、計算された実数値を切り捨てて整数化した値を n に設定する。この整数 n の値、最初に与えられた μ の値、先ほど求めた σ の値の3つ組 n, μ, σ から式を計算して g_s と g_p の数値を求めて、最初に指定していた g_s と g_p の値を修正する。そうしてあたかもそれら修正後の値が最初から指定されていたことにすれば、矛盾はなくなる。このように変更して変更を受けた値が、作成するフィルタの特性値として適切ではない場合には、 g_s と g_p の片方あるいは両方の値を変えて上記の過程を再度試行するか、もしくは μ の値を変えて試行する。(また、もしも方程式 $G(\sigma) = r$ に解 σ が存在しなかった場合には、与えた3つ組は実現が不可能であることを意味する。)

非線形方程式 $G(\sigma) = r$ の解法について

上記中の非線形方程式 $G(\sigma) = r$ は二分法を用いて以下のようにして解いた。いま $V(z) = G(z) - r$ と定義して $V(z) = 0$ を解くことにする。まず二分法の初期値を $z_L = 0.1, z_R = 10.0$ とおく (これらは求めたい解の下限と上限のつもりであり、 z_L と z_R はどちらも解ではないとする)。 $V(z_L)$ と $V(z_R)$ が同符号の場合は (適切な) 解は無いとして解法を中止する。異符号の場合は解を二分法で必ず求めることができる。そこで以下を繰り返す。

中点 $z_M \leftarrow (z_L + z_R)/2$ を取り、 $V(z_M)$ が数値的に零であるかもしくは $(z_R - z_L)/z_M \leq \epsilon_{\text{tol}}$ が成立したら解が得られたとして $z \leftarrow z_M$ として二分法を終了する。成立していない場合には、 $V(z_L)$ と $V(z_M)$ が異符号ならば $z_R \leftarrow z_M$ とし、同符号ならば $z_L \leftarrow z_M$ としてさらに繰り返しを続ける。

4 伝達関数に対応するフィルタの構成

チェビシェフ多項式を用いて表わされた $g(t)$ を伝達関数に持つフィルタ \mathcal{F} を、レゾルベントを用いて構成する方法を示す。まず σ と固有値の正規化座標 t は共に実数として

$$\frac{1}{t^2 + \sigma^2} = \frac{1}{\sigma} \operatorname{Im} \frac{1}{t - \sigma\sqrt{-1}}$$

である（ここで Im は虚部をとる操作を表わす）。さらに正規化座標 t と固有値の座標 λ の間の関係である $t = \frac{2}{b-a}(\lambda - a) - 1$ を用いると、

$$\rho \equiv \frac{a+b}{2} + \left(\frac{b-a}{2}\right)\sigma\sqrt{-1}$$

とおけば、

$$\frac{1}{t - \sigma\sqrt{-1}} = \left(\frac{b-a}{2}\right) \frac{1}{\lambda - \rho}$$

である。さらに表記の簡単化のために、

$$\ell \equiv \left(\frac{\mu^2 + \sigma^2}{\sigma}\right) \left(\frac{b-a}{2}\right)$$

とおく。すると t を用いて表わされているチェビシェフ多項式の引数の式を λ を用いて表わせば

$$2 \frac{\mu^2 + \sigma^2}{t^2 + \sigma^2} - 1 = 2\ell \operatorname{Im} \frac{1}{\lambda - \rho} - 1$$

となる。これに対応する作用素は $2\ell \operatorname{Im} \mathcal{R}(\rho) - I$ であるから、伝達関数：

$$g(t) = g_s T_n \left(2 \frac{\mu^2 + \sigma^2}{t^2 + \sigma^2} - 1 \right), \text{ あるいは } f(\lambda) = g_s T_n \left(2\ell \operatorname{Im} \frac{1}{\lambda - \rho} - 1 \right)$$

に対応するフィルタが

$$\mathcal{F} = g_s T_n (2\ell \operatorname{Im} \mathcal{R}(\rho) - I)$$

であることがわかる。ここで $\mathcal{R}(\rho)$ は虚数 ρ をシフトとするレゾルベントであり、 I は恒等作用素を表わす。

5 フィルタの作用を与える算法

上述のようにして構成された伝達関数を持つフィルタ \mathcal{F} を、与えられた実の N 次列ベクトルを m 個並べた組 X ($N \times m$ 実行列) に作用させ、別の実の N 次列ベクトルを m 個並べた組 Y (これも $N \times m$ 実行列) を作る処理 $Y \leftarrow \mathcal{F} X$ の算法は、たとえば図1のようにして実現できる。ただし V, W はそれぞれ作業用の実の列ベクトルの組 ($N \times m$ 行列) で、 Z は作業用の複素列ベクトルの組 ($N \times m$ 行列) である。またレゾルベント $\mathcal{R}(\rho)$ を作用させる処理 $Z \leftarrow \mathcal{R}(\rho) W$ は実際には与えられたベクトルの組 W からまず右辺ベクトルの組 BW を作り、次に行列 $C \equiv A - \rho B$ を係数とする連立1次方程式の組 $CZ = BW$ を Z について解くことで実現する。行列 A, B が実対称でシフト ρ は複素数なので C は複素対称行列である。連立1次方程式の右辺の組 BW は実であるが、係数行列 C や解の組 Z は複素である。行列 C を係数とする連立1次方程式の組を解く処理が全部で n 回含まれているが、 C は変化しないので、最初に1度だけ C の行列分解を計算して保持すればその後はその分解結果を用いることで新たな分解の計算は省けるので連立1次方程式の組を少ない演算量で解くことができる。

```


$$\rho \leftarrow \frac{a+b}{2} + \left(\frac{b-a}{2}\right) \sigma \sqrt{-1};$$


$$\ell \leftarrow \left(\frac{\mu^2 + \sigma^2}{\sigma}\right) \left(\frac{b-a}{2}\right);$$


$$Z \leftarrow \mathcal{R}(\rho) X;$$


$$V \leftarrow X;$$


$$W \leftarrow 2\ell \operatorname{Im} Z - X;$$

for  $k := 2$  to  $n$  do begin
    
$$Z \leftarrow \mathcal{R}(\rho) W;$$

    
$$Y \leftarrow 4\ell \operatorname{Im} Z - 2W - V;$$

    
$$V \leftarrow W;$$

    
$$W \leftarrow Y$$

end;

$$Y \leftarrow g_s Y$$


```

図 1: 中間固有値用フィルタの作用 $Y \leftarrow \mathcal{F}X$ の算法

6 伝達関数の設計の例

チェビシェフ多項式で表わされる形の伝達関数 $g(t)$ をパラメータを実際に指定して設計した例をいくつか示す。通過域は $t \in [-1, 1]$ で、阻止域は $|t| \geq \mu$ である。伝達関数の極の位置は純虚数 $t = \pm \sigma \sqrt{-1}$ である。各図のグラフでは伝達率の大きさ $|g(t)|$ の対数値を曲線で、偶関数であるので右半分 $t \geq 0$ だけをプロットした。補助線は、通過域の端 $t = 1$ とそのときの伝達率の値 g_p 、それと阻止域の端 $t = \mu$ とそのときの伝達率の値 g_s を示している。

3つ組として n, μ, σ の値を指定して伝達関数を設計した例を表 1 に示す。また、3つ組として g_p, g_s, μ を指定する場合の例を表 2 に示す（これはまず方程式を解いて σ を得て、次に n を表わす式を用いて求めた値を切り捨てて整数 n とし、そうして n, μ, σ の値から最初に指定された g_p, g_s の値を修正する）。

表 1: 3つ組として n, μ, σ の値を指定した場合の伝達関数の設計の例

n	μ	σ	g_p	g_s	グラフ
10	2.0	1.0	2.6432×10^{-4}	5.7779×10^{-13}	図 2
15	2.0	1.5	6.3773×10^{-4}	9.7139×10^{-15}	図 3
20	1.5	1.5	7.4070×10^{-6}	9.7724×10^{-16}	図 4
20	1.5	2.0	2.0812×10^{-4}	1.8190×10^{-12}	図 5
30	1.5	3.0	3.1020×10^{-4}	5.7779×10^{-13}	図 6

表 2: 3つ組として g_p, g_s, μ の値を指定した場合の伝達関数の設計の例

g_p	g_s	μ	σ	n	修正後 g_p	修正後 g_s	グラフ
10^{-4}	3×10^{-13}	1.5	1.8660	20	1.0319×10^{-4}	3.3178×10^{-13}	図 7
10^{-2}	3×10^{-13}	2.0	4.0636	32	1.0427×10^{-2}	1.3205×10^{-13}	図 8

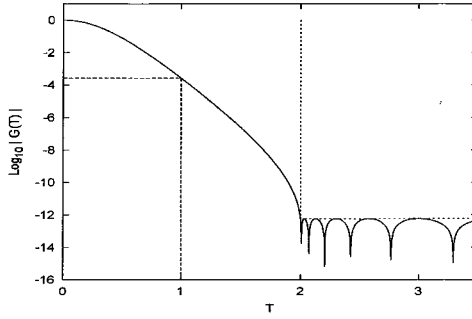


図 2: 伝達関数の大きさ $|g(t)|$ ($n = 10, \mu = 2.0, \sigma = 1.0$)

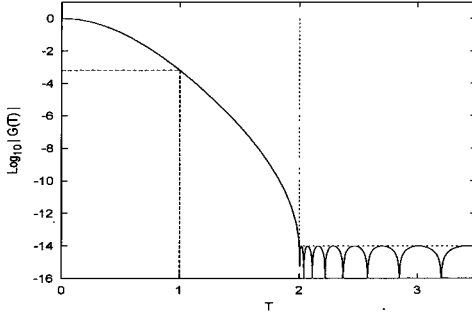


図 3: 伝達関数の大きさ $|g(t)|$ ($n = 15, \mu = 2.0, \sigma = 1.5$)

7 フィルタ対角化法の数値実験例

「シフトが虚数のレゾルベントの多項式の作用の実部」となるフィルタを、チェビシェフ多項式を利用する今回の方法で具体的に構成して、それを用いたフィルタ対角化法で中間固有値を持つ近似固有対がうまく求められるかを数値実験で調べた。以下にその実験の例を 1 つ示す。

例題の一般固有値問題（有限要素法離散化） 例題に用いた実対称定値一般固有値問題 $A\mathbf{v} = \lambda B\mathbf{v}$ は、辺の長さが π の 3 次元立方体領域における零ディリクレ境界条件での（符号が逆の）ラプラシアン $-\nabla^2$ の固有値問題を有限要素法を用いて離散化近似して得られるものである。有限要素分割は立方体領域を各辺の方向にそれぞれ 51, 61, 71 の小区間に等分割することで行なった。そうして各要素内に於ける展開基底には各辺方向の 3 重線形関数を用いた。この有限要素法の離散化で得られる行列 A と B の次数は $N = 50 \times 60 \times 70 = 210,000$ で、帯幅が小さくなるように基底関数に番号をうまく付けると各行列の半帯幅は $w_L = 1 + 50 + 50 \times 60 = 3,051$ となる。そうして固有値 λ が（固有値分布の内側に位置する）区間 $[a, b]$ に含まれる固有対を求めた。この離散化された行列の一般固有値問題の固有値は厳密値を数式に値を入れて計算できる。そうして、固有値が区間 $[170, 180]$ にある真の固有対は 81 個、固有値が区間 $[300, 310]$ にある真の固有対は 114 個、固有値が区間 $[1000, 1010]$ にある真の固有対は 180 個、であることもわかる、

使用したフィルタ 通過域 $\lambda \in [a, b]$ が $t \in [-1, 1]$ と対応する正規化座標 t で表わしたフィルタの伝達関数を、前述のように $g(t) = g_s T_n(y)$ 、ただし $y = 2x - 1$ で $x = \frac{\mu^2 + \sigma^2}{t^2 + \sigma^2}$ とする ($\mu \leq |t|$ が阻止域で、 $g(t)$ の

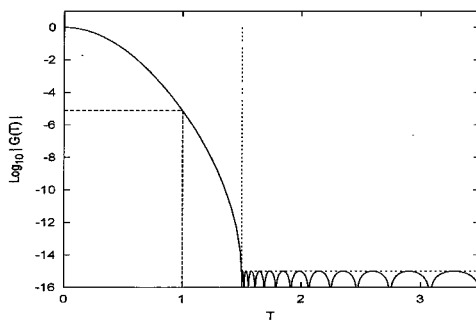


図 4: 伝達関数の大きさ $|g(t)|$ ($n = 20, \mu = 1.5, \sigma = 1.5$)

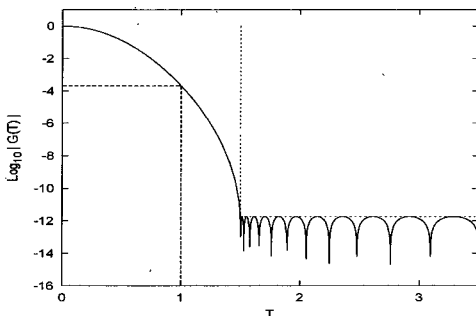


図 5: 伝達関数の大きさ $|g(t)|$ ($n = 20, \mu = 1.5, \sigma = 2.0$)

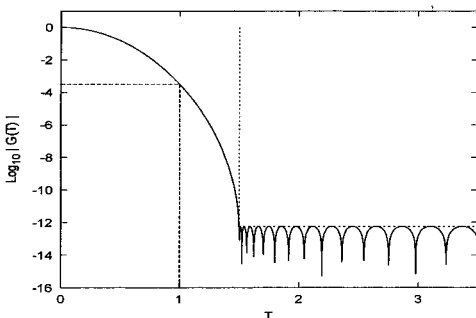


図 6: 伝達関数の大きさ $|g(t)|$ ($n = 30, \mu = 1.5, \sigma = 3.0$)

極は $t = \pm \sigma\sqrt{-1}$ となる). 以下の実験例ではすべて 3 つ組として $n=20, \mu=1.5, \sigma=2.0$ の値を指定したので, $g_p = 2.081 \times 10^{-4}$, $g_s = 1.819 \times 10^{-12}$ である (伝達関数のグラフは図 5). レゾルベントの虚数シフトは $\rho = \frac{a+b}{2} + \left(\frac{b-a}{2}\right)\sigma\sqrt{-1}$ で, また $\ell = \left(\frac{\mu^2 + \sigma^2}{\sigma}\right)\left(\frac{b-a}{2}\right)$ であるので, 区間 $[a, b]$ が $[170, 180]$ の場合には $\rho = 175 + 10\sqrt{-1}$ であり, $[300, 310]$ の場合には $\rho = 305 + 10\sqrt{-1}$ であり, $[1000, 1010]$ の場合には $\rho = 1005 + 10\sqrt{-1}$ である. またこれら 3 通りすべての場合について $\ell = 15.625$ となる. そうしてフィル

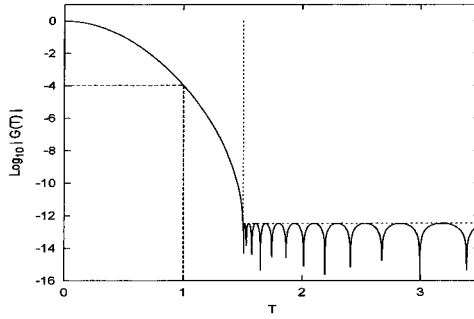


図 7: 伝達関数の大きさ $|g(t)|$ ($g_p = 1.0319 \times 10^{-4}$, $g_s = 3.3178 \times 10^{-13}$, $\mu = 1.5$)

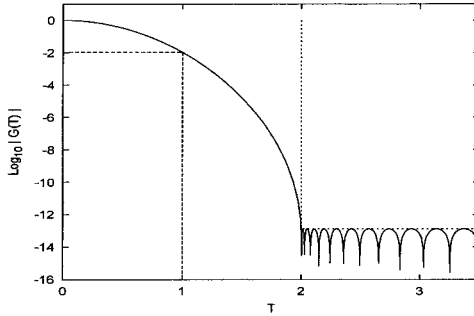


図 8: 伝達関数の大きさ $|g(t)|$ ($g_p = 1.0427 \times 10^{-2}$, $g_s = 1.3205 \times 10^{-13}$, $\mu = 2.0$)

タは $\mathcal{F} = g_s T_n (2\ell \operatorname{Im} \mathcal{R}(\rho) - I)$ と書ける。フィルタ対角化法の実験はこのフィルタを用いて行なった。

フィルタ対角化法の概略 フィルタ対角化法では、まず最初に乱数で生成された m 個のランダムなベクトルの組 ($N \times m$ 行列) を B -正規直交化して X を作成し、それにフィルタを作用させて $Y \leftarrow \mathcal{F}X$ を作る。この 2 つの組 X と Y から、フィルタの伝達特性を考慮に入れた適切な分析により、 Y の列の適切な線形結合により、区間 $[a, b]$ の近傍にある元の一般固有値問題の固有値に対応する固有空間をすべて含む不変部分空間を近似する空間の基底を構成する。レイリーリッツ法をそのように構成された基底に適用して得られたリッツ対を元の一般固有値問題の近似対にする。

残差と残差の B^{-1} -ノルムの定義 近似固有対 (λ, \mathbf{v}) に対する残差は $\mathbf{r} \equiv (A - \lambda B)\mathbf{v}$ で、残差の B^{-1} -ノルムは $\Delta \equiv \sqrt{\mathbf{r}^T B^{-1} \mathbf{r}}$ とする。ただし近似固有ベクトル \mathbf{v} はあらかじめ $\mathbf{v}^T B \mathbf{v} = 1$ と正規化されているとする。

使用計算システム CPU は Intel Core i7-5960X (8 コア) で、メモリは DDR4-2133MHz (PC4-17000) でクアドチャネル、8Gbyte のモジュールを 8 個で主記憶容量は 64Gbyte である。OS は CentOS 7 for x86_64 (64bit 版) である。コンパイラは Intel Fortran v15.0.0 for x86_64 であり、プログラムは Fortran90 と OpenMP ディレクティブでコーディングをし、コンパイラのオプションには `"-fast -openmp"` を指定して OpenMP により 8 スレッド並列で実行した。(計算中に実際に使用しているメモリの量は 21Gbyte 程度

であった。)

7.1 区間が [170, 180] で、濾過するベクトルの数が $m=150$ の場合

図 9 のグラフは、フィルタを通過する前の B -正規直交ベクトルの組 X と通過した後のベクトルの組 Y の B -内積の行列 $\beta = X^T B Y$ の固有値を減少順に並べてプロットしたものである。水平の点線は、 β の固有値の切断に用いた閾値 $10g_s = 1.819 \times 10^{-11}$ を示している。閾値で切断された β の固有値が多数あるので、ベクトルの数 $m = 150$ は既に十分である。フィルタ対角化法全体の経過時間は 847 秒であった。

図 10 のグラフは、フィルタ対角化により得られた近似対を、横軸に固有値の値をとり、縦軸に残差の B^{-1} -ノルムをとってプロットしたものである。残差のノルムはフィルタの通過域に於ける伝達率の不均一性を反映して、区間の中央付近で小さく、区間の端に向かうに従って増大して、4 桁程度の変動を示している。また、図 11 のグラフは、フィルタ対角化で得られた近似対を、横軸に固有値の値をとり、縦軸に固有値の絶対誤差の大きさをとってプロットしたものである。区間の中央部分では絶対誤差の大きさが 10^{-13} 程度なので有効精度は 15 桁程度であり、区間の端では絶対誤差の大きさが 10^{-8} 程度なので有効精度は 10 桁程度である。

7.2 区間が [300, 310] で、濾過するベクトルの数が $m=200$ の場合

図 12 のグラフは、フィルタ通過前の B -正規直交ベクトルの組 X と通過後のベクトルの組 Y の B -内積の行列 $\beta = X^T B Y$ の固有値を減少順に並べてプロットしたものである。水平の黒い点線は、 β の固有値の切断に用いた閾値 $10g_s = 1.819 \times 10^{-11}$ を示している。閾値で切断される β の固有値が多数あるので、ベクトルの数 $m = 200$ は既に十分である。フィルタ対角化法全体の経過時間は 1,009 秒であった。

図 13 のグラフは、フィルタ対角化で得られた近似対について、横軸には固有値の値をとり、縦軸には残差の B^{-1} -ノルムをとってプロットしたものである。フィルタの通過域に於ける伝達率の不均一性を反映して、残差のノルムは区間の中央付近で小さく、区間の端に向かうに従って増大して、4 桁程度の変動を示している。また、図 14 のグラフは、フィルタ対角化で得られた近似対について、横軸には固有値の値をとり、縦軸には固有値の絶対誤差の大きさをとってプロットしたものである。区間の端では固有値の絶対誤差が 10^{-8} 程度で有効精度は 10 桁程度であり、区間の中央では固有値の絶対誤差が 10^{-13} 程度で有効精度は 15 桁程度である。

7.3 区間が [1000, 1010] で、濾過するベクトルの数が $m=300$ の場合

図 15 のグラフは、フィルタ通過前の B -正規直交ベクトルの組 X と通過後のベクトルの組 Y の B -内積の行列 $\beta = X^T B Y$ の固有値を減少順に並べてプロットしたものである。水平の線は、 β の固有値の切断に用いた閾値 $10g_s = 1.819 \times 10^{-11}$ を示している。閾値で切断される β の固有値が多数あるので、ベクトルの数 $m = 300$ は既に十分である。フィルタ対角化法全体の経過時間は 1,374 秒であった。

図 16 のグラフは、フィルタ対角化で得られた近似対について、横軸には固有値の値をとり、縦軸には残差の B^{-1} -ノルムをとってプロットしたものである。フィルタの通過域に於ける伝達率の不均一性を反映して、残差のノルムは区間の中央付近で小さく、区間の端に向かうに従って増大して、約 4 桁の変動を示している。また、図 17 のグラフは、フィルタ対角化で得られた近似対について、横軸には固有値の値をとり、縦軸には固有値の絶対誤差の大きさをとってプロットしたものである。区間の端では固有値の絶対誤差の大きさは 10^{-8} 程度で、有効精度は 11 桁程度であり、区間の中央では固有値の絶対誤差の大きさは 10^{-12} 程度で有効精度は 15 桁程度である。

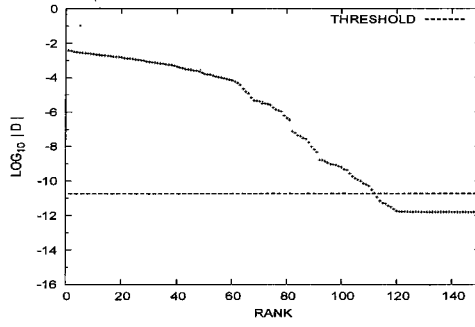


図 9: 行列 $\beta = X^T B Y$ の固有値分布 (区間 $[170, 180]$, $m=150$)

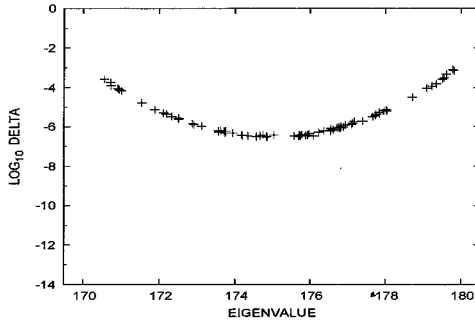


図 10: 近似対の残差のノルム Δ (区間 $[170, 180]$, $m=150$)

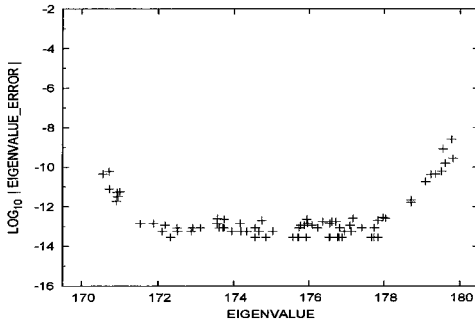


図 11: 近似固有値の絶対誤差 (区間 $[170, 180]$, $m=150$)

8 おわりに

フィルタ対角化法で実対称定値一般固有値問題の固有対で固有値が指定された区間にあるものを解くためのフィルタとして今回はレゾルベントの多項式の実部を用いる方法を考察した。用いるレゾルベントは1つだけであることと、レゾルベントの多項式は、多項式の次数に等しい回数だけレゾルベントを適用して実現されるので、複数のレゾルベントを用いる方法に比べると必要な記憶量や演算量が減らせる。

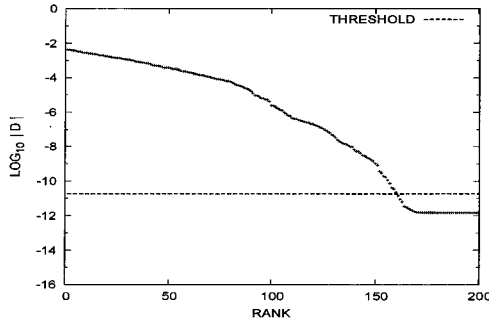


図 12: 行列 $\beta = X^T B Y$ の固有値分布 (区間 $[300, 310]$, $m=200$)

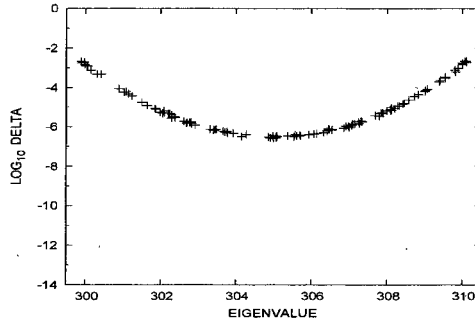


図 13: 近似対の残差のノルム Δ (区間 $[300, 310]$, $m=200$)

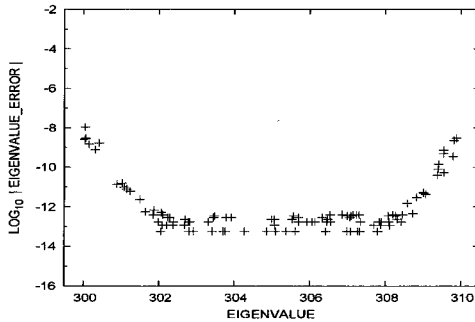


図 14: 近似固有値の絶対誤差 (区間 $[300, 310]$, $m=200$)

レゾルベントの多項式の実部をベクトルの組に作用させる計算はレゾルベントを逐次的に適用する処理になる。それに対して、フィルタとして複数のレゾルベントの線形結合を用いる方法では、ベクトルの組に各レゾルベントを作用させる処理はそれぞれ並列に処理できる。よってこのことから、レゾルベントの多項式の実部をフィルタとして採用すると、潜在的な並列性が低下すると言える。しかし扱うレゾルベントは1つだけでよく、しかもレゾルベントの作用の計算では対応する連立1次方程式を解く際に（直接法ならば）係数行列の分解、あるいは（疎行列に対する反復法ならば）係数行列の不完全分解を1度行なえばその後

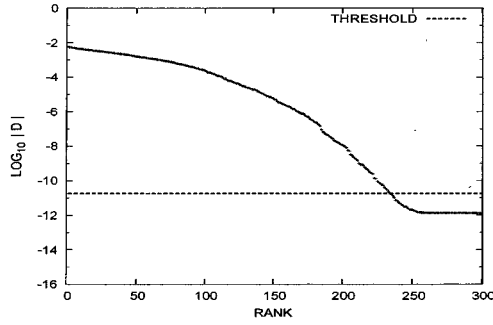


図 15: 行列 $\beta = X^T B Y$ の固有値分布 (区間 $[1000, 1010]$, $m=300$)

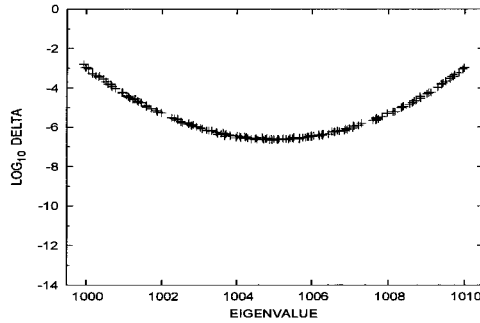


図 16: 近似対の残差のノルム Δ (区間 $[1000, 1010]$, $m=300$)

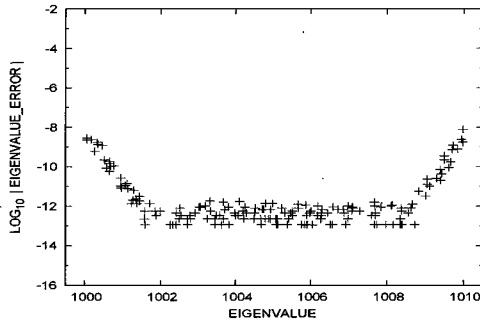


図 17: 近似固有値の絶対誤差 (区間 $[1000, 1010]$, $m=300$)

は同じ分解を再び行なう必要は無いので、合計の計算量を減らせることを期待できる。

今回はフィルタの伝達関数をチェビシェフ多項式を用いて表わされる形に制限した。それによりフィルタの設計は非常に簡単化された。フィルタの作用はチェビシェフ多項式の 3 項漸化式を利用して実現できるのも便利である。しかしシフトを 1 つだけ選んでそのレゾルベントの多項式の作用で構成されたフィルタは、シフトを複数 (8 個~16 個) 選んでそれらのレゾルベントの線形結合で構成されたフィルタと比べると、どうしてもその特性は劣るものになる。特に遷移域の幅はあまり狭くすることができない。今回の簡易

な設計法ではフィルタの伝達関数をチェビシェフ多項式を用いた式で構成した。それにより阻止域で強く一様に減衰する特性が容易に実現できたが、通過域における特性は伝達率の一様性が良くない。たとえば通過域での伝達率の最大最小比が 10^5 であれば、既にそれだけで固有ベクトルの精度を 5 桁失う可能性がある。今後は設計法を多少複雑にしてもフィルタの構成法に改良を加えて、伝達率の通過域における一様性の向上を行ないたい。

参 考 文 献

- [1] 村上弘: 固有値が指定された区間内にある固有対を解くための対称固有値問題用のフィルタの設計, **情報処理学会論文誌: コンピューティングシステム** (ACS31), Vol.3, No.3 (2010 年 9 月), pp.1-21.
- [2] 村上弘: 対称一般固有値問題のフィルタ作用素を用いた不変部分空間の近似構成, **情報処理学会論文誌: コンピューティングシステム** (ACS35), Vol.4, No.4 (2011 年 10 月), pp.1-14.
- [3] 村上弘: レゾルベントを用いたフィルタによる固有値問題の解法について, **情報処理学会研究報告**, Vol.2012-HPC-133, No.22 (2012 年 3 月), pp.1-8.
- [4] 村上弘: 実対称定値一般固有値問題の最小側固有値を持つ固有対に対する実数シフトのレゾルベントを組み合わせたフィルタによる解法, **先進的計算基盤システムシンポジウム論文集 2012**, (2012 年 5 月), pp.81-82.
- [5] 村上弘: Hermite 対称な定値一般固有値問題のフィルタ対角化法について, **情報処理学会研究報告**, Vol.2012-HPC-134, No.1 (2012 年 6 月), pp.1-8.
- [6] 村上弘: レゾルベントの線形結合をフィルタに用いたエルミート定値一般固有値問題のフィルタ対角化法, **情報処理学会論文誌: コンピューティングシステム** (ACS45), Vol.7, No.1 (2014 年 3 月), pp.57-72.
- [7] 村上弘: フィルタ対角化法について, **日本応用数理学会 2014 年度年会予稿集** (2014 年 8 月), pp.329-330.
- [8] 村上弘: レゾルベントの多項式をフィルタとして用いる対角化法について, **情報処理学会研究報告**, Vol.2014-HPC-146, No.13 (2014 年 9 月), pp.1-4.
- [9] 村上弘: 実対称定値一般固有値問題に対するレゾルベントの多項式によるフィルタの構成法の検討, **情報処理学会研究報告**, Vol.2014-HPC-147, No.2 (2014 年 12 月), pp.1-10.
- [10] 村上弘: 実数シフトのレゾルベントを組み合わせたフィルタによる実対称定値一般固有値問題の下端付近の固有値を持つ固有対の解法, **HPCS2015 シンポジウム論文集**, Vol.2015 (2015 年 5 月), pp.38-51.
- [11] 村上弘: 一つのレゾルベントから構成されたフィルタを用いた実対称定値一般固有値問題に対するフィルタ対角化法の実験, **情報処理学会研究報告**, Vol.2015-HPC-149, No.7 (2015 年 6 月), pp.1-16.
- [12] 村上弘: 実数シフトのレゾルベントの多項式をフィルタに用いた実対称定値一般固有値問題の下端付近の固有値を持つ固有対の解法, **日本応用数理学会 2015 年度年会予稿集 (統合版)** (2015 年 9 月), pp.442-443.